ESD-TR-69-184

W-7550

NONLINEAR PREDICTION AND DIGITAL SIMULATION

J. F. A. Ormsby

JUNE 1969

Prepared for

ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

Project 6151
Prepared by

THE MITRE CORPORATION
Bedford, Massachusetts

Contract AF19(628)-2390

ESD-TR-69-184

W-7550

NONLINEAR PREDICTION AND DIGITAL SIMULATION

J. F. A. Ormsby

JUNE 1969

Prepared for

ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

Project 6151
Prepared by
THE MITRE CORPORATION
Bedford, Massachusetts
Contract AF19(628)-2390

FOREWORD

The work reported in this document was performed by The MITRE Corporation, Bedford, Massachusetts, for Electronic Systems Division, Air Force Systems Command under Contract AF 19(628)-2390. This information was originally published in Working Paper W-7550, The MITRE Corporation, February 1965.

REVIEW AND APPROVAL

Publication of this technical report does not constitute Air Force approval of the report's findings or conclusions. It is published only for the exchange and stimulation of ideas.

ANTHONY P. TRUNFIO, Technical Advisor
Development Engineering Division
Directorate of Planning and Technology

ABSTRACT

An approach to digital nonlinear prediction is proposed and analyzed. The basic relations are developed. The nonlinear operator is obtained by quantization of the data.

The model is developed in terms of occupancy of data cells in N-space. Extensions to increase occupancy and reduce error are formulated. Illustrative results are included.

A comparison with linear techniques is made and over-all conclusions on error, quantization level, length of data required, time invariance, etc., are provided.

TABLE OF CONTENTS

1.0 INTRODUCTION

Methods to effect statistical prediction have involved polynomial fitting and correlation and/or spectral analysis. Both can use a minimum mean square error criterion and result in a set of weights as an optimum linear operator. In both cases, calculations of the weights based on knowledge or computation of the pertinent statistics is made.

Our concern here is with a non linear approach which involves any intermediate determination of the statistics. It also offers a readily available means for judging error and adjusting for an improved prediction. Being a non linear method, the results should be at least as good as a corresponding linear technique.

To obtain the desired non linearity, a quantization of the data is required. Of course such a quantization itself degrades the error possible with the technique. The technique as applied to digital simulation forms a variation on an approach discussed in Reference 1. We now discuss the approach.*

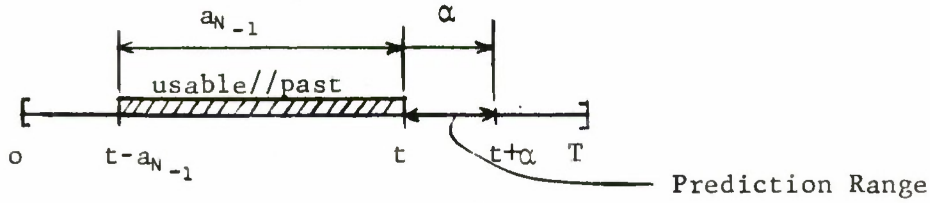
2.0 THE MODEL

We let t be the present time and T the total interval of data for processing. We let α be the time advance of prediction. Without loss of generality, we take the data interval as $[0, T]$ and consider the usable past at any t from t to $t - a_{N-1}$. For equally spaced sampled data $a_{N-1} = (N-1)\delta$ where δ is the sampling interval. These definitions may be summed up in the following sketch

---

Author's Note

*The work reported here was originally considered by the author in the summer of 1960 while at S.T.L. Other matters prevented a proper evaluation and summary at that time. The present report represents a current effort to fill this need.



Let the continuous parameter time series be $X(t)$. We then estimate $X(t+\alpha)$ by

$X^*(t+\alpha)$, the predicted value at $t+\alpha$ using $X(\sigma)$; $\sigma \leq t$

In the sampled finite past case we have

$X^*_N(t+\alpha)$, the predicted value at $t+\alpha$ using $X(t-a_i)$; $i=o, 1, \cdots, N-1$

i.e. by $\{X_o, X_{-1}, X_{-2} \cdots, X_{-(N-1)}\}$.

Although $\Delta_i a = a_{i+1} - a_i$ may not be equal, nothing is lost by assuming so, since the average value of the Δa is constrained by the sampling theorem. We take

$X^*(t+\alpha) = F[X(t), X(t-\delta), \cdots, X(t-a_{N-1})]$ where

F is a time dependent non linear function. In this case we can always make the instantaneous square error $\epsilon = \left(X(t+\alpha) - F[X(t), X(t-a_1), \cdots, X(t-a_{N-1})]\right)^2 = o.$

However since F is to be used when $X(t+\alpha)$ is not available, it is better not to have F time dependent. Hence we consider a range of times over which a time invariant operator F must minimize the mean squared error given by

$$\overline{\epsilon_\alpha^2} = \frac{1}{\beta_\alpha} \int_{a_{N-1}}^{T-\alpha} (X(t+\alpha) - F[X(t), X(t-a_1), \cdots, X(t-a_{N-1})])^2 dt$$

with $[a_{N-1}, T-\alpha]$ the maximum range over which to consider error or in other words over which we can distinguish a "present" value. In the above $\beta_\alpha = T - \alpha - a_{N-1}$. As seen the value of $X^*(t+\alpha)$ depends on the number, N, of data points used in the memory.

We replace † F by an infinite series of terms whose orthogonality is invariant to X. Thus,

$$F = \sum_{-\infty}^{\infty} An\ \Phi_n[X(t),\ X(t-a_1),\ \ldots,\ X(t-a_{N-1})]$$

where, independent of X values,

$$\frac{1}{T-\alpha-a_{N-1}} = \int_{a_{N-1}}^{T-\alpha} \Phi_n[\ ]\ \Phi_m[\ ]\ dt = 1;\ n = m$$
$$= 0;\ n \neq m$$

Then minimum $\overline{\epsilon_\alpha^{\ 2}}$ gives, using $\frac{\partial \epsilon\alpha^2}{\partial A_n} = 0$ and the orthogonality of the $\{A_n\}$

$\{\Phi\}$,

$$A_n = \frac{1}{T-\alpha-a_{N-1}} \int_{a_{N-1}}^{T-\alpha} X(t+\alpha)\ \Phi_n[\ ]\ dt$$

$$= \frac{\overline{X\Phi n}}{\overline{\Phi_n^{\ 2}}} = \frac{< X,\ \Phi_n >}{< \Phi_n,\ \Phi_n >}$$

It is impossible, however, to have completeness using such a representation on any X(t) (i.e., for continuous valued X). However, quantization of X provides a realization of the desired orthogonal set for all X(t). In other words quantizing X allows construction of a $\{\Phi_n\}$ set for all X, the approximation being dependent on the fineness (degree) of quantization. As X is more finely quantized, the

$$F[X(t),\ \ldots\ ,\ X(t-a_{N-1})] \rightarrow \Sigma A_n\ \Phi_n\ [X(t),\ \ldots\ ,\ X(t-a_{N-1})]$$

for any X(t) with the $\{\Phi_n\}$ orthogonal and independent of X.

† See Reference 1

The orthogonality depends however on the range β_α, chosen. We then have

$$\overline{\epsilon_\alpha^{2}} = \frac{1}{\beta\alpha}\int_{a_{N-1}}^{T-\alpha} (x(t+\alpha) - \sum_{\infty}^{\infty} A_n \Phi_n [\;])^2 dt$$

$$= \frac{1}{\beta\alpha}\left(\int_{a_{N-1}}^{T-\alpha} X^2(t+\alpha)dt - \sum_{0}^{\infty} \beta_n A_n^2\right)$$

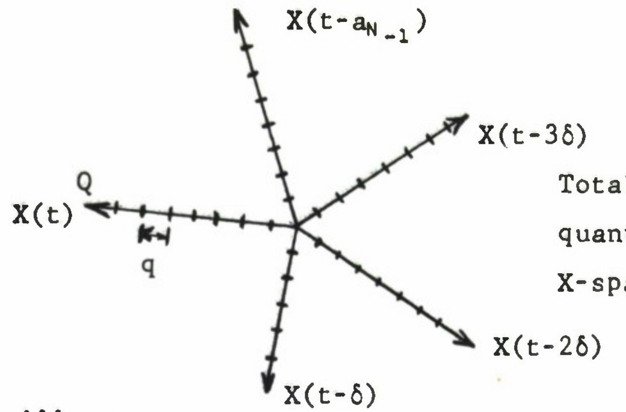
$$= \frac{1}{\beta_\alpha}\left(\int_{T-\alpha-\beta_\alpha}^{T-\alpha} X^2\ (t+\alpha)dt - \sum_{0}^{\infty} \frac{\left(\int_{T-\alpha-\beta\alpha}^{T-\alpha} X(t+\alpha)\Phi_n\ dt\right)^2}{\beta n}\right)$$

with $$\beta n = \int_{T-\alpha-\beta\alpha}^{T-\alpha} \Phi n^2\ dt$$

We see that A_n represents the projection of predicted X values in the Φ_n direction of the $\{\Phi_n\}$ space. Thus $\overline{\epsilon_\alpha^2}$ compares the true value $\dfrac{\int_{T-\alpha-\beta\alpha}^{T-\alpha} X^2(t+\alpha)dt}{\beta_\alpha}$ with $\sum_0^\infty \left(\dfrac{\beta n}{\beta\alpha}\right) A_n^2$ where $\dfrac{\beta n}{\beta_\alpha}$ represents the probability of getting A_n over β_α. For a fixed t, $X(t+\alpha)$ is estimated by a single A_n for some Φ_n.

3.0 QUANTIZATION and REALIZATION

If for each n, the $\Phi_n\ [X(t), X(t-\alpha_1), \ldots, X(t-\alpha_{N-1})] = 1$ or 0 as a function of the N - dimensional argument of X values, then $A_n = \dfrac{[X,\Phi n]}{\beta n}$ is a conditional expectation. That is the average of $X(t+\alpha)$ conditioned on the occurrence of $\Phi_n(t) = 1$ as t varies over β_α and β_n counts the number of such occurrences. Let us now consider for equally spaced data the N - dimensional space of $\{X(t-k\delta)\}_0^{N-1}$ values. Let each $X(t-k\delta)$ range be divided into Q quantum each of width q as shown.



Total number of quantum cells in X-space is Q^N.

Let $n = 1, 2, 3, \ldots, Q^N$

We take $\Phi_n [\{X(t-i\delta)\}_o^{N-1}] = 1 \quad$ if $\{X\}_o^{N-1} \in$ cell n

$= 0 \quad$ if $\{X\}_o^{N-1} \notin$ cell n

Thus if Φ_n occurs at t then we take A_n as $X^*(t+\alpha)$ and A_n is calculated as average of $X(t'+\alpha)$ values at all $t' < t$ times when Φ_n occurs. The $\{\Phi_n\}$ set remains fixed only if the cell structure in N-space does. We note that the orthogonality of the $\{\Phi_n\}$ is independent of the t range of integration. Also since $< \Phi_n,\Phi_n > = 0$ only if $\Phi_n = 0$, the

$$A_n = \frac{< X, \Phi n >}{< \Phi n,\Phi n >} \text{ are bounded}$$

The normality condition

$$\frac{1}{\beta\alpha}\int_{T-\alpha-\beta_\alpha}^{T-\alpha} \Phi_n^* (\ ) \, dt = \frac{1}{\beta\alpha}\int_{T-\alpha-\beta\alpha}^{T-\alpha} \Phi_n \, dt$$

depends on the X values that is which member of the ensemble is chosen and the span of calculation $\beta\alpha$. We see that $\left\{\frac{\Phi n}{\beta^{\frac{1}{2}}_n}\right\}$ produces an orthonormal set; $\beta_n = [\Phi_n,\Phi_n]$.

It is useful to rewrite the formulae for A_n and $\overline{\epsilon_\alpha^2}$ in a style which reflects the cell matching imposed by the quantization as reflected by the $\{\Phi_n\}$ set.

$$A_n = \frac{\int_{\tau_n} X\ (t+\alpha)\,dt}{\int_{\tau_n} dt} = \frac{\int_{\tau_n} X\ (t_{i,n}+\alpha)\,dt}{\tau_n}$$

where $\tau_n = \{t_i\colon \Phi n\ [t] = 1,\ t \in [a_{N-1}\ ,\ T-\alpha]\}$ so that

$A_n = \overline{X(t+\alpha)}$ conditioned an occurrence of cell n.

$$= \frac{\sum_{i=1}^{Nn} X(t_i+\alpha)}{N_n} \quad \text{for sampled data}$$

and $\sum_{n=1}^{Q^N} N_n = \beta_\alpha$ for a fixed cell structure over $t_i \in [a_{N-1}\ ,\ T-\alpha]$ so that any $\{X\}_o^{N-1}$ sequence can be in no more than one cell.

For non-multiple use of data sets $\{X\}_o^{N-1}$ but allowing overlapping cells the total number of matches is still β_α. However the orthoganality of the Φ_n set and the expressions for A_n and $\overline{\epsilon_\alpha^{\ 2}}$ previously developed rely on the cells in N-space being non-overlapping.

We also have

$$X^*(t+\alpha) = \Sigma A_n \Phi_n = \{A_n : \Phi_n(t) = 1,\ t \in [a_{N-1}\ , T-\alpha]\} \quad 1 \le n \le Q^N$$

and

$$\overline{\epsilon_\alpha^{\ 2}} = \frac{\sum_{i=1}^{\beta_\alpha} [X(t_i + \alpha) - A_{n_i}]^2}{\beta_\alpha}$$

Incidentally we may view the quantization as a means to set up a special set of orthoganal functions for any X by writing Φ as a product given by

$$\Phi_{n,m,k,\cdots,\ell}[X(t),X(t-a_1);\cdots,X(t-a_{N-1})]=\Phi_n[X(t)]\Phi_m[X(t-a_1)]\ldots$$

Finally $\epsilon_\alpha^2 = (X(t+\alpha) - F[\ ]^2 = (X(t+\alpha) - \Sigma A_n\Phi_n)^2$

which can be written $(X(t+\alpha) - \Sigma A_n\Phi_n - (F[\ ] - \Sigma A_n\Phi_n))^2$

$$\text{or} \quad (X(t+\alpha) - F - (\Sigma A_n\Phi_n - F))^2$$

These forms indicate two kinds of errors.

(1) Optimum F is not sufficient, i.e. future not predictable from the past (not usual).

(2) $\Sigma A_n\Phi_n$ is not a good enough approximation, i.e. orthoganal set not complete.

For Φ_n derived via quantization, the level q plays a roll in (2) and indeed contributes a quantization error as well.

Finally we remark that when $F = F(t)$ that is the Φ_n are changing with t, the number of possible A_n = number of possible cells $> Q^N$.

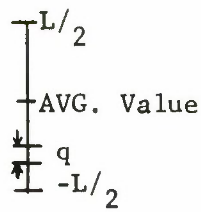
3.1 Quantization Error and Data Calibration

Given a joint distribution over data, we can always construct an orthonormal function set relative to the distribution. For example, if X is normally distributed, we choose the Φ_n as Hermite polynomials. In dealing with time series we interchange time and ensemble averages with time averages taken over sufficiently long data stretches providing the process is ergotic and so stationary. With X normal, the optimum predictor is a linear one. This result can serve as a reference to the non-linear models.

Approximating F via quantization gives

$$A_n = < X, \Phi_n; L >$$

where L is the peak to peak signal variation in the interval used. We choose the quantum step q so as to produce the desired ϵ^2 and divide the X range into $Q = L/q$ quantum steps.



Quantization thus converts X(t) into a step function S(t) where any step level S_i results from X values given by $S_i - q/2 \leq X < S_i + q/2$. The quantization error, e, is a random variable with the X distribution and

$-q/2 \geq e < q/2$. With q small enough we assume that e is approximately uniformly distributed in any one interval so that the phase (ensemble) average of e^2 is given by

$$\text{AVG.}(e^2) = \frac{1}{q}\int_{-q/2}^{q/2} e^2 de = \frac{2}{q}\,\frac{q^3}{3x8} = q^2/12$$

$$= \frac{1}{12}\left(\frac{L}{Q}\right)^2$$

If we assume 1024 divisions for + or - values (a representative value for some A-D converters) then Q = 2048 and

$$\text{AVG.}(e^2) \sim \left(\frac{L}{2^{10}}\right)$$

Thus writing the mean square error of quantization as K q^2, K, a constant, the mean square error in the prediction model $\overline{\epsilon^2} \geq K q^2$

From

$$\overline{\epsilon_\alpha^2} = \frac{1}{\beta_\alpha} \int_{T-\alpha-\beta\alpha}^{T-\alpha} X^2(t+\alpha)\,dt - \sum_{o}^{Q} \beta_n A_n^2$$

if quantization level q is too large, $A_{no} \rightarrow \overline{X}$ for some single $n=n_o$

and $A_n \rightarrow o$, $n \leq n_n$, $\beta_{no} \rightarrow \beta_\alpha$,

$$\overline{\epsilon^2} \sim 0 \left(\frac{1}{\beta_\alpha} \int_{T-\alpha-\beta\alpha}^{T-\alpha} (X(t+\alpha) - \overline{X})^2 \quad dt\right) \sim 0(K'q^2) \quad , K' > K.$$

Of course $\overline{\epsilon^2}$ may be large with q small for operator F not near optimum.

It is well known that in general the non-linear operator extends bandwidth due to intermodulation of components. Also, if we consider quantization error in N-space as a distance d, then

$$d_{max} = \sqrt{\sum_{i=1}^{N} q_i^2}$$ with q_i the quantum level in the i^{th} dimension.

If $q_i = q$ for all i,

$$d_{max} = (Nq^2)^{\frac{1}{2}} = N^{\frac{1}{2}}q$$

For preserving constant distance (error), we have

$$q \sim \frac{1}{N^{\frac{1}{2}}}$$

so that the more past samples used to form a data set $\{X\}^{N-1}$, the smaller q required to maintain the quantization error.

With α, β_α, N, q variable there is a range of possibilities with various influences on $\overline{\epsilon_\alpha^2}$.

As N increases, q should decrease. With decreasing q and increasing

N the β_n, number of matches over β_α, should drop. This means that the number of matches of data sets $\{X(t'\}_o^{N-1}$ to some reference set $\{X(t)\}_o^{N-1}$, $t' < t$ decreases. Thus although quantization error may be fixed, variability of the estimate of an A_n (associated with $\{X(t)_o^{N-1}\}$ may rise due to the reduced number of matches. In general we desire as many cell matches as possible at each t.

For example, unless q is large (and so larger prediction errors) or a trend sensitive parameter, such as conditional expectation function, is used, no reasonable prediction can be made on trend type data since obtaining cell matches becomes difficult or impossible.

One method to increase number of cell occupancies involves using multiple use of data and/or overlapping cell structure with time. These procedures allow for time varying $\{\Phi_n\}$ sets and form a degression from the theory presented above. Details of implementing such procedures are discussed in the next section.

Another way to improve cell matching is to deal with essentially trend free data using polynomial interpolation techniques. Simple linear calibration may be sufficient.

The calibration of the input data can be given as

$$X(t) = ay(t)+b = L(y)$$

where L represents a linear operation $a(\neq o)$ and b are known factors which can be time varying.

In the prediction model we operate an $X(\sigma)$, $\sigma \leq t$ with F to give $X^*(t+\alpha)$ so that in a 1:1 manner we have

$$y^*(t+\alpha) = \frac{X^*(t+\alpha)-b}{a}$$

4.0 METHODS OF ESTABLISHING CELL OCCUPANCY

As noted above we desire models which increase cell occupancy. This involves basically

1. defining the cell structure
2. establishing occurrence of occupancy

We now discuss three variations on dealing with these two factors.

1) This definition establishes from the next new data set left from the $\{X\}_o^{N-1}$ data sets remaining after having filled all previous cells set up. This allows no multiple use of data and reduced amount of cell overlap. A data set defining a cell is taken as the center of the cell. In order to maximize the number of cell prediction the data sets are used starting first with the data set associated with the current time, t. The total number of sets equals β_α.

2) This is a variation on (1) which allows multiple use of any set $\{X\}_o^{N-1}$ for matching except those previously used as cell centers. Thus both multiple use of data and cell overlap occurs. We have K(number of cells setup) $\leq \beta_\alpha \leq \sum_{j=1}^{K} d_{nj}$ = total number of sets used in the K setup cells where d_{nj} = number of matching $\{X\}_o^{N-1}$ sets into cell setup at n_j.

If we could essentially assume no cell overlap that is a fixed $\{\Phi n\}$ set (fixed cell structure) then for (1) and (2) we would have

$$\overline{\epsilon_\alpha^2} = \frac{1}{K} \sum_{j=1}^{K} (X(t_{i_j}+\alpha)-A_{i_j})^2$$

$$A_{i_j} = \frac{1}{d_{ij}(\alpha)} \sum_{ij=1}^{d_{ij}(\alpha)} X(t_{ij}+\alpha)$$

3) Another variation takes all past sets as cell centers so that $K = \beta_\alpha$. This method gives more cell overlap and multiple use of data but maximizes cell occupancy at each t in $[a_{N-1}, T-\alpha]$ and places the center of the cell for t at $\{X(t)\}_0^{N-1}$ at each t. Briefly in summary (1) tends to minimum occupancy (overlap) with the maximum of these at current time t while (3) produces maximum occupancy (overlap). We call Method A, Approach 3) and Method B, Approach 1) together with cells held non-overlapping. We first discuss Method B for $\overline{\epsilon_\alpha^2}$ and A_n. We also include another variation called Model C.

4.1 Method B

With fixed cell structure that is $\{\Phi_n\}$ set fixed, we have from before,

$$\overline{\epsilon_\alpha^2} = \frac{1}{\beta_\alpha}\left(\int_{T-\alpha-\beta_\alpha}^{T-\alpha} X^2(t+\alpha)dt - \sum_0^\infty \frac{\left(\int_{T-\alpha-\beta_\alpha}^{T-\alpha} X(t+\alpha)\Phi_n [\ ]dt\right)^2}{\beta_n}\right)$$

with sampled (discrete time series) data

$$\overline{\epsilon_\alpha^2} = \frac{1}{\tilde{\beta}_\alpha} \sum_{k=1}^{K} \left(\sum_{j=1}^{n_k(\alpha)} (X(t_{jk}+\alpha) - A_{jk})^2\right); \quad \tilde{\beta}_\alpha = \sum_{k=1}^{K} n_k(\alpha)$$

with
$$A_{n_k} = \frac{1/\tilde{\beta}_\alpha \sum_{j=1}^{n_k} X(t_{jk}+\alpha)\Phi_{jk}}{1/\tilde{\beta}_\alpha \sum_{j=1}^{n_k} \Phi_{jk}^2} = \frac{\sum_{j=1}^{n_k} X(t_{jk}+\alpha)}{n_k}$$

so that
$$\overline{\epsilon_\alpha^2} = \frac{1}{\tilde{\beta}_\alpha}\left(\sum_{n=1}^{\tilde{\beta}_\alpha} X^2(t_n+\alpha) - \sum_{k=1}^{K} n_k {Aj_k}^2\right)$$

$$= \frac{1}{\tilde{\beta}_\alpha}\left(\sum_{n=1}^{\tilde{\beta}_\alpha} X^2(t_n+\alpha) - \sum_{k=1}^{K} \frac{\left[\sum_{j=1}^{n_k} X(t_{jk}+\alpha)\right]^2}{n_k}\right)$$

The matchings (for each α) to the cell setup by past set $\{X(T)\}_0^{N-1}$ taken at present time T(i.e., n_T) produce A_{n_T} as $X^*(T+\alpha)$ for each α.

4.2 Method A

As noted, the method sets up each of the $\tilde{\beta}_\alpha$ data sets as centers for cells. The approach thus allows multiple use of any data set $\{X\}_0^{N-1}$ and overlapping cell structures.

Now,

$$\overline{\epsilon_\alpha^2} = \frac{1}{\tilde{\beta}_\alpha}\sum_{n=1}^{\tilde{\beta}_\alpha}[X(t_n+\alpha)-A_n]^2 = \frac{1}{\tilde{\beta}_\alpha}\sum_{n=1}^{\tilde{\beta}_\alpha}[X^2(t_n+\alpha)-2X(t_n+\alpha)A_n+A_n^2]$$

for each of the $\tilde{\beta}_\alpha$, $n=1,2,\ldots,\tilde{\beta}_\alpha$ where we take

$$A_n = \frac{1}{\ell_n(\alpha)}\sum_{j=1}^{\ell_n(\alpha)} X(t_{jn}+\alpha) \quad ; \quad \sum_{n=1}^{\tilde{\beta}_\alpha} \ell_n \geq \tilde{\beta}_\alpha$$

Then, without orthogonality effects allowed since the $\{\Phi_n\}$ sets correspond to overlapping cells,

$$\overline{\epsilon_\alpha^2} = \frac{1}{\tilde{\beta}_\alpha}\sum^{\tilde{\beta}_\alpha}\left[X^2(t_n+\alpha)-2X(t_n+\alpha)\frac{n}{\ell_n}\sum_{j=1}^{\ell_n} X(t_{jn}+\alpha) + \left(\frac{\sum_{j=1}^{\ell_n} X(t_{jn}+\alpha)}{\ell_n}\right)^2\right]$$

4.3 Model C

In order to increase further the number of occupancies and so hopefully reduce $\overline{\varepsilon_\alpha^2}$ averaged over α for α min < a < a max, for the cell specified at the current T value, we consider sequentially $\{X(T-j\delta)\}_0^{N-1}$ as potentially equivalent $\{X(T)\}_0^{N-1}$ data sets with adjusted α values†. This is valuable when $\{X(t)\}_0^{N-1}$ has none or few matches. For α the data sets then used for equivalent prediction points are those at $t \le T-\alpha$. We designate the corresponding data sample numbers as n_T and n_{T_j} where $t_j = T-\alpha-j\delta$ where δ is the data spacing.

Thus, at α using the set at n_{T_j}, we use $\alpha_j = \alpha+j\delta$, $j = 0,1,\ldots,R$ and look for occupancy of the cell determined by $\{X(T-j\delta\}_0^{N-1}$ by the sets $\{X(t)\}_0^{N-1}$, $t < T - j\delta$ that is for $n < n_{T_j}$.

At j the maximum α value possible as in previous methods is determined by $\alpha_{max_j} = n_{T_j} - n$ where n are the matching data sets $\{X\}_0^{N-1}$ to $\{X(T-j\delta)\}_0^{N-1}$. The minimum α in $\alpha_{min_j} = j\delta$.

We take

$$\varepsilon_j \triangleq \frac{1}{\alpha_{max_j}} \sum_{\alpha j=\alpha_{min_j}}^{\alpha_{max_j}} \overline{\varepsilon_{\alpha j}^2} \quad \text{for each } j.$$

Then we choose the n_{T_j} and associated matching sets which give $\min_j \varepsilon_j$ and calculate the A_n set for the $\alpha_{min_j} < \alpha_j < \alpha_{max_j}$ corresponding to the α. If $\alpha_{max_j} < \alpha_{max}$, then for $n_{T_j} < n_{T_j}$ we choose the next lowest ε_j which has $\alpha_{max_j} > \alpha_{max_j}$. The process is repeated until $\alpha_{max_j} = \alpha_{max}$ or the j range is depleted.

† Of course increased occupancy must be weighed against basically higher $\overline{\varepsilon_\alpha^2}$ as α increases.

4.4 Adaptive Control

As we have seen, parameters N, q, β_α influence $\overline{\epsilon_\alpha^{\ 2}}$. It is thus useful to have criterion, even if only crude, by which to judge such effects.

We have noted the error amplitude resolution in the sampling process is $q_o^{\ 2}/12$ where q_o is the amplitude resolution or imposed quantization range q. For example, with binary data and $q = 2^k$, the error is $4^{k-1}/3$.

Referenced to the original (unquantized) data before sampling and excluding other error sources, the total error $\overline{\epsilon^2_{TOT}}$ is approximately $\overline{\epsilon_\alpha^{\ 2}} + q^2/12$ with

$$o \le \overline{\epsilon_\alpha^{\ 2}} \le \frac{1}{\beta_\alpha} \int_{T-\alpha-\beta_\alpha}^{T-\alpha} X^2(t+\alpha)\,dt \sim \frac{1}{\tilde{\beta}_\alpha} \sum_{n=1}^{\tilde{\beta}_\alpha} X^2(t_n+\alpha)$$

Thus, we may usually expect

$$q^2/12 \le \overline{\epsilon^2_{TOT}} \le q^2/12 + \frac{1}{\tilde{\beta}_\alpha} \sum_{n=1}^{\tilde{\beta}_\alpha} X^2(t_n+\alpha)$$

With $\overline{\epsilon_\alpha^{\ 2}} = \frac{1}{\tilde{\beta}_\alpha} \sum_{n=1}^{\tilde{\beta}_\alpha} X^2(t_n+\alpha) - \sum_{0}^{Q} \beta_k A_k^{\ 2}$ and $q^2/12$ not truly independent errors since $\overline{\epsilon_\alpha^{\ 2}}$ depends on q, we might have $\overline{\epsilon^2_{TOT}} \lessgtr \overline{\epsilon^2_\alpha}$ for example.

From

$$\overline{\epsilon_\alpha^{\ 2}} = \overline{X^2(t+\alpha)} - \sum_{k=1}^{K} \frac{\left[\sum_{j=1}^{n} X(t_{jk}+\alpha)\right]^2}{n_k}$$

as $n_k \to o$ (no cell matching) $\overline{\epsilon_\alpha{}^2} \to \overline{X^2(t+\alpha)}$.

We may then consider using the values $q^2/2$ and $\overline{X^2(t+\alpha)}$ as bounds for testing $\overline{\epsilon_\alpha{}^2}$ when changing q, N, β_α to determine acceptability or for improving $\overline{\epsilon_\alpha{}^2}$ to obtain $X^*(T+\alpha)$.

4.5 Population of Cell Occupancies

With Q steps of magnitude q and N past point data sets there are Q^N possible cells (i.e., A_n's using fixed structure). The percentage actually used is much smaller, say less than 1% on finite sections of data. In digital simulation it is not required to store all the possible or even expected $\{A_n\}$ set.

For example, with N=2 and independent normally identically distributed data at X_n, X_{n-1} for all n with a standard deviation of σ the number of A_n drops from Q^2 to $3\sigma(2Qq-3\sigma)$, $qQ/2 > 3\sigma$. An estimate of 3σ can be taken as the largest amplitude increment between successive samples.

Runs have been made on data to determine and remove trends using data types such as

(1) monotonic increasing data of a component of velocity sampled at 10 times per second.

(2) periodic high frequency type data from accelerometer residuals sampled at 10 times per second.

(3) radar trajectory position data at 10 times per second.

A definite $X^*(t+\alpha)$ versus q relationship was difficult to establish for various ranges of N and β_α.

4.6 Comparison of Methods A and B

It is convenient to display errors for comparison by

$$\% \text{ rms error} = \% \left(\overline{\epsilon_\alpha^2} \right)^{\frac{1}{2}} = \sqrt{\frac{\overline{\epsilon_\alpha^2}}{\overline{x^2}}}$$

and % actual error = % $\epsilon_\alpha^* = \frac{X^* - X}{X}$

Due to the flexibility of centering cells, it is expected that A gives better X^* (low % ϵ_α^*) than B so that the possibility of cell occupancy is more likely.

It is recalled that ϵ_α^* was taken with respect to prediction ($x^* = A_{n_T}$) point sample n_T only while $\overline{\epsilon_\alpha^2}$ in Method A was calculated over matches not only to data set for n_T but matches to data sets for all n used.

It is suggested a revised method A, say A^1, be examined to remedy these situations by (1) allowing actual predictions to be calculated over the data range and not just at n_T and (2) calculating $\overline{\epsilon_\alpha^2}$ at each prediction point, n_T', from matches of past data sets to the data set at n_T' only. At each n_T', $\{X(T')\}_0^{N-1}$ is considered fixed allowing

$A_n = \frac{\langle X, \Phi_n \rangle}{\langle \Phi_n, \Phi_n \rangle}$ to be used with cell center fixed by $\{X(T')\}_0^{N-1}$ which shifts with T' to make $\{A_n\}$ set time varying with n_T'.

5.0 GENERAL LINEAR PREDICTOR

"General" is taken to mean no assumption of stationarity. The use of the general linear predictor provides a basis of comparison for the non linear predictor.

For continuous (nonquantized) data the form of the basic system of equations for determining the coefficients of the optimum linear predictor (operator) are

$$\int_{o}^{T-\alpha} X(t-\tau_n)\, X(t+\alpha)dt = \sum_{m=o}^{N-1} \left(\int_{o}^{T-\alpha} X(t-\tau_n)\, X(t-\tau_m)dt \right) a_m$$

$$n=0,1,\ldots,N-1$$

which for sampled data becomes (with a slight abuse of notation)

$$\sum_{k=o}^{\tilde{\beta}_\alpha} X_{k-n}\, X_{k+\alpha} = \sum_{m=o}^{N-1} \left(\sum_{k=o}^{\tilde{\beta}_\alpha} X_{k-n}\, X_{k-m} \right) a_m;\ n=0.1,\ldots,N-1$$

It can be noted that an analogous form of equation appropriate to the nonlinear method is given by

$$\sum_{k=o}^{\tilde{\beta}_\alpha} \Phi_n\, X_{k+\alpha} = \sum_{m=o}^{M} \left[\sum_{k=o}^{\tilde{\beta}_\alpha} \Phi_n(X_k, \ldots)\, \Phi_m(X_k, \ldots) \right] A_m;\ n=0,1,\ldots,M$$

If we define $R(\xi,\eta) \triangleq \sum_{k=o}^{\tilde{\beta}_\alpha} X_{k-\xi} X_{k-\eta}$, we have

$$\underset{(nX1)}{\{R(n,-\alpha)\}} = \underset{(nXm)}{((R(n,m)))}\ \underset{(mX1)}{\{a_m\}}$$

so that

$$\{a_m\} = ((R(n,m)))^{-1}\ \{R(n,-\alpha)\}$$

with $((R(n,m)))$ and $\{R(n,-\alpha)\}$ both functions of α and the point from which prediction is made.

With allowance for variable α, a comparison with the nonlinear method indicates that for the same amount of information, linear processing time is higher by a factor as high as 10.

We remark that calculations at data rates higher than that provided can be afforded in all processing modes by a sliding Lagrangian fit such as a cubic.

Finally comparisons are for all prediction points having one or more matches in the nonlinear processing. Also, the rms values of actual errors taken over these points are also found useful.

6.0 OBSERVATIONS, REMARKS AND CONCLUSIONS

It was found useful to use $\overline{\varepsilon_0}$ = r.m.s. value of observed error, $\overline{\varepsilon_\alpha^2}^{\frac{1}{2}}$ calculated over matches associated with a given point of prediction

and $\hat{\varepsilon}_A$ = r.m.s. value of actual error, $\left(\overline{(\varepsilon^*_A)^2}\right)^{\frac{1}{2}}$ averaged over all x : x available as point of prediction is changed.

6.1 Linear

For small α , the average $\hat{\varepsilon}_A$ increases with N as should be expected since the linear operator becomes potentially more limited as the extent of the past (memory) increases. The error $\overline{\varepsilon_0}$ increases by about a 1.6 factor as α goes from 1 to 2.

6.2 Nonlinear

For probability of occupancy as low as 5%, the $\hat{\varepsilon}_A$ remain reasonably fixed. For q low enough, $\overline{\varepsilon_0}$ decreases with increased N as the occupancy probability drops. In other words, at each prediction point the $\overline{\epsilon_o}$ decreasing measures the degree of correct selectivity. As N increases with q high enough, large variation in actual error $X^* - X$ is allowed. Indeed $\hat{\epsilon}_A$ decreases with increasing N while $\hat{\epsilon}_A$ increases by a factor of 1.5 as α goes from 1 to 2. As q decreases the $\hat{\epsilon}_A$ and $\overline{\epsilon_o}$ agree. As α increases the q value necessary for this agreement increases.

For low N and α (say 1), $\hat{\epsilon}_A$ can decrease with β_α while $\overline{\epsilon_o}$ increases for the same q. Both $\hat{\epsilon}_\alpha$ and $\overline{\epsilon_o}$ reduce with increased fs (sampling rate) for the same effective β_α, N, α. Also, $\hat{\epsilon}_A$ and $\overline{\epsilon_o}$ appear relatively insensitive to the interval between prediction points or to the over-all span of these points, the latter effect relying on the degree of stationarity present.

6.3 Linear Versus Nonlinear (for a given data sampling rate)

Although the linear procedure requires no quantization, comparisons with the nonlinear method are more equitable with the nonlinear results when this discrepancy is taken into account. Basically unless a sufficiently low q can be used the $\hat{\epsilon}_A$ values from linear processing are lower as was the case for α = 1,2 and all the N, β_α and data used. As q lowers to favor the nonlinear error, the probability of a prediction occurring falls sharply for the β_α range used.

The sample conditional expectation obtained with finite length records and the lack of completeness with q values > o result in gaps in the determination of the conditional expectation on which, indeed, the prediction is based.

A first and crude step toward overcoming this deficiency can be taken via a two point interpolation as follows. (More exactly as N-dimensional interpolation on the data sets $\{X(t)\}_0^{N-1}$ of nearest matches could be made to specify $X^*(t+\alpha)$ associated with $\{X(T)\}_0^{N-1}$).

We let n be the count corresponding to time t as before and $\{X\}_{n_1}$ and $\{X\}_{n_2}$ be the two N-dimensional data sets having the two smallest distances d_1 and d_2 ($\leq d_1$) respectively in N-space from $\{X\}_{n_T}$. We take

$$X^*(T+\alpha) = X(t_1+\alpha) + \frac{d_1}{d_1+d_2} \Delta X;$$

with

$$\Delta X = X(t_2+\alpha) - X(t_1+\alpha)$$

The minimum mean square error criterion results in the predicted value $X^*(t+\alpha)$ being given by a quantized version (for determining cell occupancy) of the sample conditional expectation $E(X(t+\alpha)|C)$ where C is the cell whose center is determined by $\{X(T)\}_0^{N-1}$ rather than the non-quantized version $E(X(t+\alpha) \mid \{X(T)\}_0^{N-1})$. We have

$$X^*(t+\alpha) = E(X(t+\alpha) \mid C) = \int_X X(t+\alpha)\ P(X(t+\alpha) \mid C)\ dx$$

and for sampled data

$$X^*(t+\alpha) = \Sigma X\ P\alpha(X|C)$$

With a finite stretch of data we must consider

$$P\alpha(X_k \leq X < X_{k+1}) \mid C) = \frac{K_{\alpha,k,c}}{\ell_{\alpha,c}}$$

where $K_{\alpha,k,c}$ = number of times an $X(t+\alpha)$ value from the $\tilde{\beta}\alpha$ sequence of X values lies within the (X_k, X_{k+1}) range given that its corresponding past data set $\{X(t)\}_0^{N-1}$ enters cell C.

$\ell_{\alpha,c}$ = number of past data sets which enter cell C over the $\tilde{\beta}\alpha$ sequence of data points.

We may note here the various first order, joint and conditional probabilities as follows:

Shortened Notation

$$P_\alpha[X_k \leq X < X_{k+1} \mid C] = \frac{K_{\alpha,k,c}}{\ell_{\alpha,c}} = P_\alpha(X|C)$$

$$P_\alpha(C) = \frac{\ell_{\alpha,c}}{\tilde{\beta}_\alpha}$$

$$P[X_k \leq X < X_{k+1}, C] = \frac{K_{\alpha,k,c}}{\tilde{\beta}_\alpha} = P_\alpha(C)\ P_\alpha(X|C) = P_\alpha(X,C)$$

$$P[X_k \leq X < X_{k+1}] = \frac{K_{\alpha,k}}{\tilde{\beta}_\alpha} = P_\alpha(X)$$

$$P[C|X_k \leq X < X_{k+1}] = \frac{K_{\alpha,k,c}}{K_{\alpha,k}} = \frac{P_\alpha(X,C)}{P_\alpha(X)} = P_\alpha(C|X)$$

where $K_{\alpha,k,c}$ and $\ell_{\alpha,c}$ are defined above and $K_{\alpha,k}$ is the number of times an X value in the $\tilde{\beta}_\alpha$ sequence of data points lies within the range (X_k, X_{k+1}).

In Method A', we have over a range of prediction points

$$\sum_c P_\alpha\ [X_k \leq X < X_{k+1}, C] \neq P_\alpha(X_k \leq X < X_{k+1})$$

since the cells overlap. Indeed

$$(\sum_c P_\alpha[X_k \le X < X_{k+1}\ , C] - P_\alpha(X_k \le X < X_{k+1}) = \frac{(\sum_c K_{\alpha,k,c}) - K_{\alpha,k}}{\mathfrak{F}_\alpha}$$

However,

$$\Sigma_k\ P_\alpha(X_k \le X < X_{k+1},\ C) = P_\alpha(C)$$

Another approach, of course, for determining $X^*(t+\alpha)$ could be based on the maximum likelihood estimator, this is, on MAX $P_\alpha(X(t+\alpha) \mid C)$ where C is the cell determined by $\{X(T)\}_0^{N-1}$, then $X^*(t+\alpha)$ can be taken as $X_{k_{m+\frac{1}{2}}}$ where k_m produces the maximum value of $K_{\alpha,k,c}$.

REFERENCE

Bose, A. G., "A Theory of Nonlinear Systems," M.I.T. Research Laboratory of Electronics Technical Report 309, May 15, 1956

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* The MITRE Corporation Bedford, Massachusetts	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

NONLINEAR PREDICTION AND DIGITAL SIMULATION

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

N/A

5. AUTHOR(S) *(First name, middle initial, last name)*

J. F. A. Ormsby

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
June 1969	28	1

8a. CONTRACT OR GRANT NO. AF 19(628)-2390 b. PROJECT NO. 6151 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) ESD-TR-69-184 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* W-7550

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
N/A	Electronic Systems Division, Air Force Systems Command, L. G. Hanscom Field, Bedford, Massachusetts

13. ABSTRACT

An approach to digital nonlinear prediction is proposed and analyzed. The basic relations are developed. The nonlinear operator is obtained by quantization of the data.

The model is developed in terms of occupancy of data cells in N-space. Extensions to increase occupancy and reduce error are formulated. Illustrative results are included.

A comparison with linear techniques is made and over-all conclusions on error, quantization level, length of data required, time invariance, etc., are provided.

DD FORM 1 NOV 65 1473

Security Classification

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Nonlinear Prediction						
Digital Simulation						
Data Quantization Sampling Effects						

Security Classification